1. **Descriptive Analytics**

DataSet-1: Historic health dataset

It contains historic health care data for heart risk.

Here’s a brief description of the fields.

17 features-

|  |  |
| --- | --- |
| Varibles | Description |
| P\_ID | Patient Unique ID |
| male | Gender |
| Age | age |
| Education | Education code |
| currentSmoker | Smoker or Not |
| cigsPerDay | Cigrate per Day |
| BPMeds | Having BP Medicines |
| prevalentStroke | Stroke history |
| prevalentHyp | Hypertension history == HBP |
| diabetes | Diabetes |
| totChol | Total cholesterol |
| sysBP | Sys BP> |
| diaBP | Dia BP< |
| BMI | BMI == Body mass index |
| heartRate | Heart rate reading |
| glucose | Glucose reading |
| HeartRisk | Heart Risk Target Variable |

**Task**: Analyze this data and come up with insights from the data. You can choose any tool/programming language. You are free to choose the measures that you want to focus on.

**Output**: You can present the result of your analysis in a word doc/ppt/pdf/excel or html page. Also attach the code/steps that were used.

1. **Predictive Analytics**

DataSet-1: Historic health dataset –Use it for model Building

Dataset-2: Historic health dataset – Test dataset

**Task**: Build a model using DataSet-1 and use this model to predict heart risk for Dataset-2 (test dataset). Select the best model to predict heart risk.

**Output**: Model script and a csv or excel file with only 2 columns (details are mentioned below)

* PID – Patient ID from test dataset
* HeartRisk – Predicted value

**Tools**: R or python will be preferred.