Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

Department of Computer Science and Engineering (Data Science)

NAME: **HARSH VIKAS SHETYE** BATCH: **A2** SAP ID: **60009210068**
COURSE CODE: **DJ19DSL402** AY: **2022-23** DATE: **16 / 04 / 2023**
COURSE NAME: **Machine Learning - I Laboratory** CLASS: **S.Y.B.Tech (A)**

Mini Project
Task 3
(Model)

Case Study Title: Predictive Analytics for Insurance Claims

Aim: The problem statement for "Predictive Analytics for Insurance" involves developing a predictive model that can accurately assess the likelihood of an insurance claim being filed, as well as the potential cost of settling the claim. The goal of this project is to reduce the time and resources required for claims processing and improve the accuracy of claim assessments.

Name of the dataset: Insurance_data2.csv

Source of dataset: Kaggle Dataset Link
Why this dataset:
This dataset contains information on the insurance claim. Each observation is different policyholder with various features like the age of the person, the gender of the policyholder, body mass index, providing an understanding of the body, number of children of the policyholder, smoking state of the policyholder and individual medical costs billed by health insurance.

Feature	Description
age	age of policyholder
sex	male(1)/female(0)
bmi	body mass index(kg / m2)
children	number of children/dependents of policyholder
smoker	smoking state nonsmoker(0)/smoker(1)
region	residential area northeast(0)/northwest(1)/southeast(2)/southwest(3)
charges	medical cost

insuranceclaim	yes(1)/no(0)

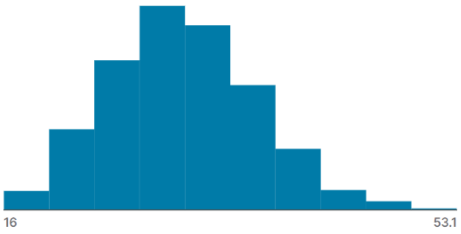
age

Valid	1338	100%
Mismatched	0	0%
Missing	0	0%
Mean	39.2	
Std. Deviation	14	
Quantiles	18	Min
	27	25%
	39	50%
	51	75%
	64	Max

sex

Valid	1338	100%
Mismatched	0	0%
Missing	0	0%
Mean	0.51	
Std. Deviation	0.5	
Quantiles	0	Min
	0	25%
	1	50%
	1	75%
	1	Max

bmi

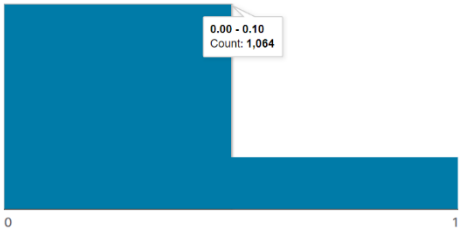


Valid	1338	100%
Mismatched	0	0%
Missing	0	0%
Mean	30.7	
Std. Deviation	6.1	
Quantiles	16	Min
	26.3	25%
	30.4	50%
	34.7	75%
	53.1	Max

children

Valid	1338	100%
Mismatched	0	0%
Missing	0	0%
Mean	1.09	
Std. Deviation	1.21	
Quantiles	0	Min
	0	25%
	1	50%

smoker

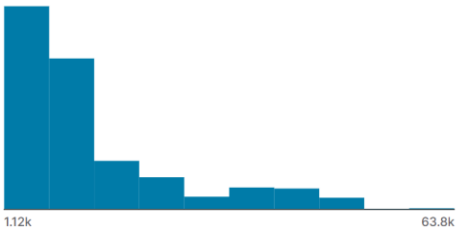


Valid	1338	100%
Mismatched	0	0%
Missing	0	0%
Mean	0.2	
Std. Deviation	0.4	
Quantiles	0	Min
	0	25%
	0	50%
	0	75%
	1	Max

region

Valid	1338	100%
Mismatched	0	0%
Missing	0	0%
Mean	1.52	
Std. Deviation	1.1	
Quantiles	0	Min
	1	25%
	2	50%
	2	75%
	3	Max

charges



Valid	1338	100%
Mismatched	0	0%
Missing	0	0%
Mean	13.3k	
Std. Deviation	12.1k	
Quantiles	1.12k	Min
	4.74k	25%
	9.39k	50%
	16.7k	75%
	63.8k	Max

insuranceclaim

Valid	1338	100%
Mismatched	0	0%
Missing	0	0%
Mean	0.59	
Std. Deviation	0.49	
Quantiles	0	Min
	0	25%
	1	50%
	1	75%
	1	Max