Data Cleaning Process Summary

Summary of the cleaning steps performed on the dataset:
#1: Initial Dataset Glimpse
#2: Replacing implausible zero values with NA for specific columns.
#3: Applying median imputation for missing values.
#4: Converting incorrect data types.
#5: Categorizing 'Age' into age groups.
#6: Labeling Outcome as Non−Diabetic (0) and Diabetic (1).
#7: Cleaned dataset saved to /data/processed/cleaned_data.csv

Summary Table of Cleaned Data − Part 1

Var1	Var2	Freq
	Pregnancies	Min. : 0.000
	Pregnancies	1st Qu.: 1.000
	Pregnancies	Median : 3.000
	Pregnancies	Mean : 3.845
	Pregnancies	3rd Qu.: 6.000
	Pregnancies	Max. :17.000
	Glucose	Min. : 44.00
	Glucose	1st Qu.: 99.75
	Glucose	Median :117.00
	Glucose	Mean :121.66
	Glucose	3rd Qu.:140.25
	Glucose	Max. :199.00
	BloodPressure	Min. : 24.00
	BloodPressure	1st Qu.: 64.00
	BloodPressure	Median : 72.00
	BloodPressure	Mean : 72.39
	BloodPressure	3rd Qu.: 80.00
	BloodPressure	Max. :122.00
	SkinThickness	Min. : 7.00
	SkinThickness	1st Qu.:25.00

Summary Table of Cleaned Data − Part 2

Var1	Var2	Freq
	SkinThickness	Median :29.00
	SkinThickness	Mean :29.11
	SkinThickness	3rd Qu.:32.00
	SkinThickness	Max. :99.00
	Insulin	Min. : 14.0
	Insulin	1st Qu.:121.5
	Insulin	Median :125.0
	Insulin	Mean :140.7
	Insulin	3rd Qu.:127.2
	Insulin	Max. :846.0
	BMI	Min. :18.20
	BMI	1st Qu.:27.50
	BMI	Median :32.30
	BMI	Mean :32.46
	BMI	3rd Qu.:36.60
	BMI	Max. :67.10
	DiabetesPedigreeFunction	Min. :0.0780
	DiabetesPedigreeFunction	1st Qu.:0.2437
	DiabetesPedigreeFunction	Median :0.3725
	DiabetesPedigreeFunction	Mean :0.4719

Summary Table of Cleaned Data − Part 3

Var1	Var2	Freq
	DiabetesPedigreeFunction	3rd Qu.:0.6262
	DiabetesPedigreeFunction	Max. :2.4200
	Age	Min. :21.00
	Age	1st Qu.:24.00
	Age	Median :29.00
	Age	Mean :33.24
	Age	3rd Qu.:41.00
	Age	Max. :81.00
	Outcome	Non−Diabetic:500
	Outcome	Diabetic :268
	AgeGroup	Length:768
	AgeGroup	Class :character
	AgeGroup	Mode :character