Assignment:-

Total Marks: 10

Submission Format: Jupyter Notebook (.ipynb)

Submission Procedure: Upload the notebook file on a new Github Repository and share the link on the QOBE portal assignment activity's student notes.

Scenario:

You are working as a data scientist for a healthcare analytics company. You have been provided with a dataset containing anonymized patient data. Your task is to clean the data, visualize it, analyze statistical relationships, and simulate a basic health-related system using First Order ODEs.

Q1: Outlier Detection and Feature Reduction (2 Marks)

- Use VarianceThreshold to remove constant features (if any).
- Detect and remove outliers from CholesterolLevel using any method of choice.
- Replace any missing values.

Q2: Data Visualization (2 Marks)

- Generate a box plot comparing BloodPressure between different RiskCategory levels.
- Create a count plot showing the number of patients in each RiskCategory per Region.

Q3: Statistical Hypothesis Testing (3 Marks)

Perform the following:

- Is the average CholesterolLevel significantly different from 200 mg/dL having a sample size of 20?
- Is there a significant relationship between Region and *RiskCategory*?

Include:

- Null and alternative hypotheses
- Test statistic, p-value
- Your interpretation (reject/fail to reject H_0)

Q4: Modeling a Health System with ODE (3 Marks)

Model the progression of a patient's medication in the bloodstream.
Equation:

$$dC/dt = k(D - C)$$

Where:
- $k = 0.1$, $D = 100$, $C(0) = 0$
- Solve this ODE for $t = 0$ to 50 using both odeint and solve_ivp
- Plot the result and explain the difference in output between the two solvers, if any.