Group 12- dataset exploration: Framingham heart study

1. Research questions :

a. Novel questions: -

- What are the key risk factors influencing the 10-year risk of coronary heart disease (chd)?
- How can these risk factors be effectively utilized to enhance early intervention strategies?
- In what ways do gender-based susceptibility patterns influence the development of heart disease, and are there gender-specific risk factors that should be addressed in targeted preventive strategies?

b. Potential audience: - Healthcare professionals, policymakers, and researchers in the field of cardiovascular health.

c. Interest to audience: - Stakeholders would be interested in understanding and predicting chd risks for early intervention, reducing the impact of heart diseases on public health.

2. Data:

a. Datasource: - Kaggle dataset from the framingham heart study (link: framingham_dataset). Over 4,000 records with 15 attributes, including demographic, behavioral, and medical factor.

b. Data period: - The time frame covered by the dataset is not explicitly mentioned.

c. Level of observation: - Observations are at the individual level, with each row representing a unique patient.

d. Repeated observations: - There are no repeated observations for a given individual.

3. Model :

a. Outcome of interest (y variable): - The outcome variable is the 10-year risk of chd (binary: "1" for "yes, " "0" for "no").

b. Predictors (x variables): - sex, age, current smoker, cigs per day, bp meds, prevalent stroke, prevalent hyp, diabetes, tot chol, sys bp, dia bp, bmi, heart rate, glucose.

c. Statistical model(s): - Logistic regression is selected due to its appropriateness for binary outcomes, facilitating the development of a predictive model that estimates the probability of chd based on the identified predictors

4. Results/conclusions :

a. Final results and conclusions:

The project aims not only to identify significant risk factors for chd but also to provide a practical and applicable logistic regression model for predicting the 10-year risk.

b. Expected conclusions:

Early analyses might reveal gender-based susceptibility, age-related influences, and potential correlations between behavioral and medical factors and the risk of chd.