Questions

**Demographic Analysis:**

* What is the average age of males and females in the dataset?
* What is the distribution of education levels across different age groups?

**Health Risk Analysis:**

* How does the smoking status correlate with the risk of CHD (Coronary Heart Disease)?
* What is the average BMI of individuals with diabetes compared to those without diabetes?

**Chronic Condition Analysis:**

* What is the prevalence of hypertension (prevalenthyp) among males and females?
* How does the presence of diabetes affect the average systolic blood pressure (sysbp)?

**Lifestyle Analysis:**

* Is there a significant difference in cholesterol levels (totchol) between smokers and non-smokers?
* What is the relationship between the number of cigarettes smoked per day (cigsperday) and heart rate?

**Risk Factor Analysis:**

* How does the prevalence of stroke (prevalentst) differ between individuals with different educational backgrounds?
* What is the average glucose level in individuals at high risk of CHD compared to those not at risk?

**Which demographic group (by age, sex, and education level) has the highest average cholesterol levels?**

* Analyze average total cholesterol levels across different age groups, genders, and education levels.

**What is the prevalence of diabetes among individuals with different education levels?**

* Determine the percentage of individuals with diabetes across different education levels.

**How does the average number of cigarettes smoked per day vary between different age groups and genders?**

* Calculate and compare the average number of cigarettes smoked per day among various age groups and between males and females.

**How does heart rate vary with BMI categories?**

* Investigate the distribution of heart rate across different BMI categories (e.g., underweight, normal weight, overweight, obese).

**How does smoking status affect the risk of developing CHD?**

* Compare the incidence of chdrisk between smokers and non-smokers.

**Questions for second database**

1. **What is the distribution of Coronary Heart Disease (CHD) risk by age group and gender?**

- Visualize the chdrisk across different age groups for both males and females.

2. **How does smoking status affect the risk of developing CHD?**

- Compare the incidence of chdrisk between smokers and non-smokers

3. **Is there a correlation between the number of cigarettes smoked per day and the risk of CHD?**

- Analyze the relationship between cigsperday and chdrisk.

4. **How does the presence of high blood pressure (prevalenthyp) relate to the risk of CHD?**

- Examine the proportion of chdrisk among individuals with and without prevalenthyp.

**5. What is the impact of diabetes on the likelihood of developing CHD?**

- Compare chdrisk between diabetic and non-diabetic individuals.

6. **How do cholesterol levels (totchol) vary among individuals with and without CHD risk?**

- Plot the distribution of totchol for those with and without chdrisk.

7. **What is the relationship between BMI and the risk of CHD?**

- Investigate how different BMI ranges correlate with chdrisk.

8. **Does the use of blood pressure medication (bpmeds) influence CHD risk?**

- Compare chdrisk among those who take bpmeds and those who do not.

9. **How do systolic (sysbp) and diastolic (diabp) blood pressures correlate with CHD risk**

- Analyze the relationship between both sysbp and diabp with chdrisk.

10. **What are the average heart rates and glucose levels for individuals with and without CHD risk?**

- Compare the mean heartrate and glucose levels for those with and without chdrisk.

**KPIs**

1. No of Patients
2. Average BMI
3. Total No of Patients
4. Average age
5. No of female patients
6. No of male patients
7. Average Systolic blood pressure(sysbp)
8. Average total cholesterol level

**About the dataset**

**Sex:** This column represents the gender of the individuals (female- male).

**Age:** This column represents the age of the individuals in the dataset. Age is a crucial factor in assessing the risk of coronary heart disease.

**Education:** This column represents the level of education of the individuals. It could be coded using categorical values indicating different levels of education attainment.

**Smokingstatus:** This column likely represents the smoking status of the individuals, indicating whether they are smokers(yes), non-smokers(no).

**Cigsperday:** If an individual is a smoker, this column represents the number of cigarettes smoked per day.

**Bpmeds:** This column indicates whether the individual is taking blood pressure medications (binary: 0 for not taking, 1 for taking).

**Prevalentstroke:** This column indicates whether an individual has had a stroke prior to the study (binary: 0 for no, 1 for yes).

**Prevalenthyp:** This column indicates whether an individual has hypertension (binary: 0 for no, 1 for yes).

**Diabetes:** This column indicates whether an individual has diabetes (binary: 0 for no, 1 for yes).

**Totchol:** This column represents the total cholesterol level of the individuals.

**Sysbp:** This column represents the systolic blood pressure of the individuals.

**Diabp:** This column represents the diastolic blood pressure of the individuals.

**BMI:** This column represents the Body Mass Index (BMI) of the individuals, which is a measure of body fat based on height and weight.

**Heartrate:** This column represents the resting heart rate of the individuals.

**Glucose:** This column represents the fasting blood glucose level of the individuals.

**Chdrisk:** This column likely represents the Ten-Year Coronary Heart Disease (CHD) Risk for each individual, which is the target variable that you may want to predict or analyze.