Summary and Recommendation

Project Overview: The goal of this analysis was to explore various health-related factors and predict the likelihood of disease risk based on demographic, clinical, and behavioral data. The dataset consists of key variables such as age, gender, smoking behavior, blood pressure, cholesterol, BMI, and medical history. The primary objective was to identify patterns and relationships between these factors and disease risk (`risco_doenca`), and to build a predictive model to assess future health outcomes.

Key Findings:

1. **Demographics and Risk Correlation:**
    - **Age and Gender Impact**: Our analysis revealed that certain age groups showed higher disease risk. Specifically, older individuals were more likely to fall into the high-risk category. Gender also played a role, with males exhibiting a higher disease risk in some cases.
    - **Education Level**: The educational background was linked to smoking habits and other health behaviors, influencing risk factors like cholesterol and blood pressure.
2. **Behavioral Factors:**
    - **Smoking**: Smokers had a higher risk of developing health issues, particularly those with a high number of cigarettes smoked per day. This was strongly reflected in the cholesterol and blood pressure data.
    - **Physical Activity and Obesity**: BMI (Body Mass Index) was another significant factor. Individuals with higher BMIs (obesity) showed a greater likelihood of having hypertension, high cholesterol, and diabetes.
3. **Medical History:**
    - **Hypertension and Diabetes**: A history of hypertension and diabetes was a strong predictor of high disease risk, as seen in individuals who were already on blood pressure medication or who had been diagnosed with these conditions in the past.
    - **Cholesterol Levels**: Elevated cholesterol levels were strongly correlated with increased disease risk, reinforcing the importance of regular monitoring and intervention for cardiovascular health.

Predictive Modeling:

To predict disease risk, a **Logistic Regression** model was built using demographic, behavioral, and clinical features. The model successfully identified high-risk individuals and demonstrated strong accuracy in classification.

- **Model Performance**: The model achieved a high level of accuracy and precision, with the ROC curve showing a favorable trade-off between sensitivity and specificity.

- **Important Features**: The most influential features for predicting disease risk included **cholesterol levels**, **blood pressure**, and **BMI**. These variables provided significant insights into identifying individuals who might be at risk for cardiovascular diseases and other health complications.

Visualization Insights:

- **Correlation Heatmap**: A heatmap revealed strong correlations between variables like cholesterol, blood pressure, and disease risk, validating the clinical importance of these factors.
- **Risk Distribution**: Boxplots and bar charts effectively showed how different variables, such as smoking and BMI, influenced disease risk. Smokers with high BMI had particularly elevated cholesterol and blood pressure levels, further supporting the need for early interventions.
- **Age vs. Risk**: The relationship between age and disease risk was clearly visualized, with older age groups showing a significantly higher proportion of high-risk individuals.

Recommendations:

Based on the findings, the following recommendations can be made:

- **Health Interventions**: Focus on educating individuals about the dangers of smoking, maintaining a healthy BMI, and managing cholesterol and blood pressure through lifestyle changes and medication.
- **Targeted Health Screening**: Prioritize screening for individuals with high BMI, elevated cholesterol, and a history of hypertension, particularly in older populations.
- **Preventive Measures**: Encourage regular check-ups and early detection of conditions like hypertension and diabetes to mitigate long-term health risks.

Conclusion:

This analysis has successfully demonstrated the critical relationships between demographic, behavioral, and clinical factors and disease risk. By understanding these factors and predicting risk accurately, healthcare providers can take a more proactive approach in managing public health, improving patient outcomes, and reducing the burden of chronic diseases. The predictive model built in this project serves as a valuable tool for early detection and targeted health interventions.