**Final Report: Analysis of Heart Attack Risk Factors**

**1. Executive Summary**

This report details the analysis of a clinical dataset of 302 patients to identify the key predictive factors for heart disease. The exploratory data analysis revealed that Maximum Heart Rate (thalach), ST Depression induced by exercise (oldpeak), and the type of Chest Pain (cp) are the most significant clinical indicators. A predictive model was developed using a Tuned Random Forest Classifier, which achieved an accuracy of **76.3%** and, more importantly, a **recall of 88%** for correctly identifying patients with heart disease. The findings were consolidated into an interactive Tableau dashboard to provide clinicians with a powerful tool for dynamic risk assessment. The results strongly suggest that these identified factors can serve as a robust foundation for an effective, data-driven system for predicting cardiovascular events.

**2. Data Cleaning and Preparation**

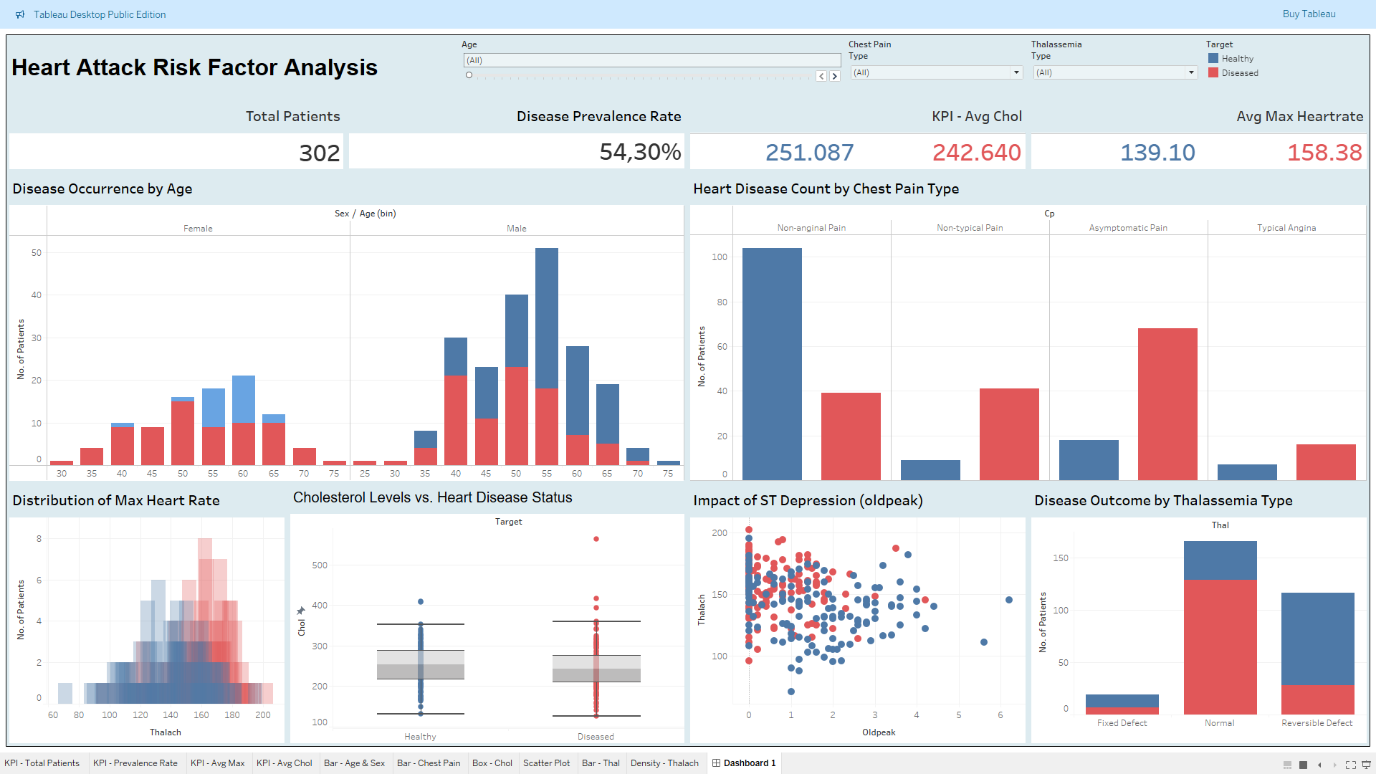
The initial dataset contained 303 patient records and 14 clinical variables. A preliminary inspection revealed one duplicate record, which was subsequently removed, resulting in a final dataset of 302 unique entries. There were no missing values, ensuring a clean and complete dataset for analysis. To improve clarity and usability in visualizations, numerical categorical variables (e.g., target, sex, cp) were converted to dimensions and assigned descriptive aliases (e.g., changing target values of 0 and 1 to "Healthy" and "Diseased," respectively).

**3. Exploratory Data Analysis (EDA) & Key Insights**

An interactive Tableau dashboard was created to visualize the complex relationships within the data. The following key insights were derived from this analysis.

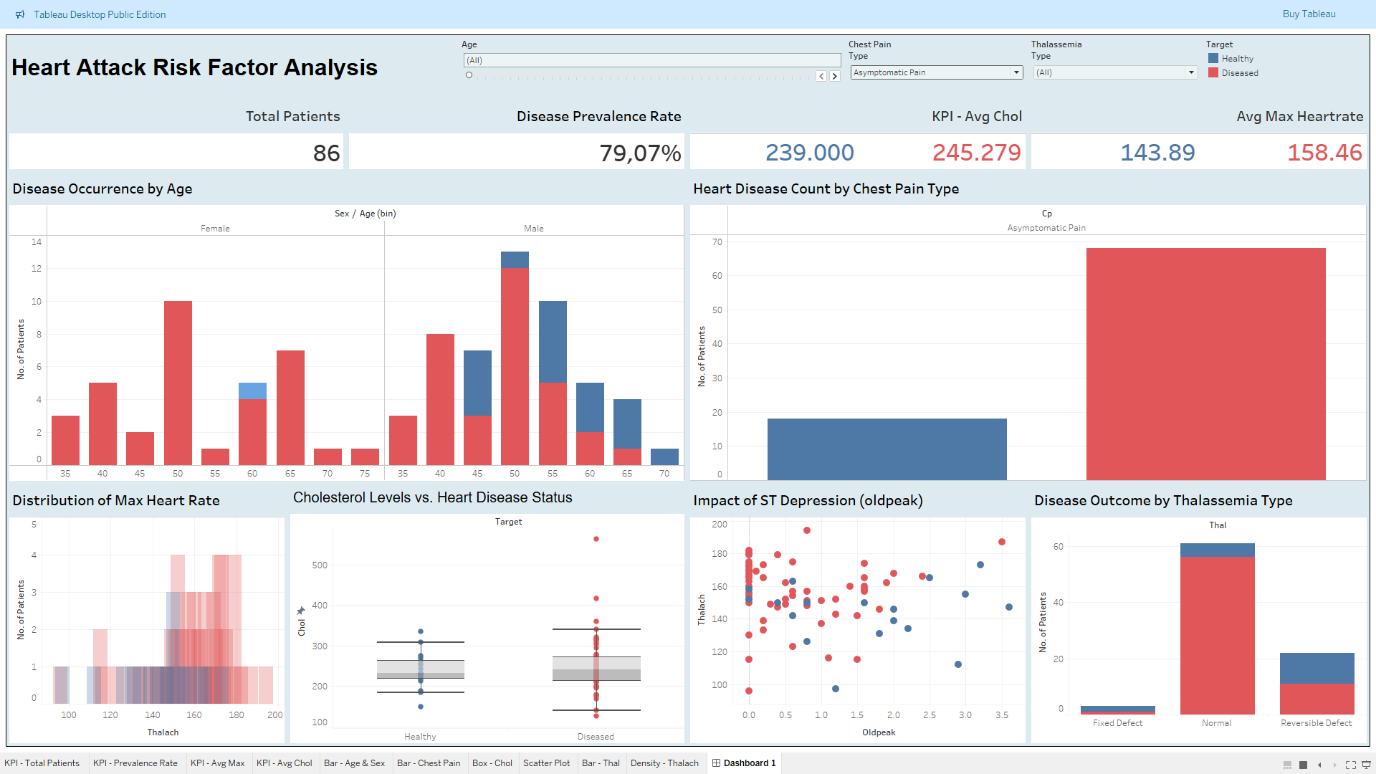
**Insight 1: Disease Prevalence and Demographic Profile**

The overall disease prevalence in the patient cohort is **54.3%**. The risk of heart disease is not uniform across demographics. As shown in the dashboard, the incidence of heart disease begins to rise significantly after the age of 40 and peaks in the 55-65 age group. While males constitute a larger portion of the overall patient population (68.2%), the prevalence of the disease is high across both genders within these high-risk age brackets.

**  
**Figure 1:** Overview of Patient Demographics and Key Clinical Indicators.

**Insight 2: Symptomatic vs. Asymptomatic Chest Pain is a Critical Indicator**

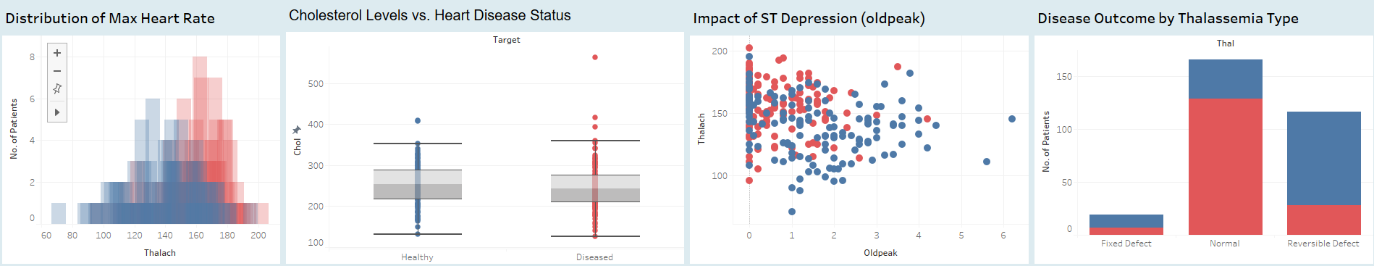
The type of chest pain (cp) reported by a patient is a powerful predictor of heart disease. The dashboard clearly shows that patients reporting **"Asymptomatic" chest pain (Type 2)** have the highest count of confirmed heart disease cases, significantly more than those with other pain types. This counterintuitive finding is critical, as it suggests that the absence of typical pain does not imply the absence of risk.

**  
**Figure 2:** Analysis showing the high incidence of heart disease among patients with asymptomatic chest pain.

**Insight 3: Key Clinical Markers Show Clear Separation**

Two clinical metrics stand out as strong differentiators between healthy and diseased patients:

* **Maximum Heart Rate (thalach):** The healthy patient group consistently achieved a higher maximum heart rate (average of 158 bpm) compared to the diseased group (average of 139 bpm). The overlapping density plot below visualizes this clear separation.
* **ST Depression (oldpeak):** The scatter plot shows that a higher oldpeak value, especially when combined with exercise-induced angina (exang = 1), is strongly correlated with the presence of heart disease.

**  
**Figure 3:** Key clinical indicators showing a clear distinction between the two patient groups.

**4. Predictive Modeling and Evaluation**

To quantify the predictive power of these factors, a baseline Logistic Regression model and a more advanced **Tuned Random Forest Classifier** were developed. The Random Forest model, after hyperparameter tuning, provided superior performance.

The final model was evaluated on a test set of 76 patients. The results are summarized in the classification report and confusion matrix below.

*(Insert the image of your "Final Model - Confusion Matrix (Tuned Random Forest)" here)*  
**Figure 4:** Confusion Matrix for the final predictive model.

*(Insert the final "Classification Report" from your notebook here)*  
**Table 1:** Classification Report for the final model.

The model achieved an overall accuracy of **76.3%**. Most critically for a clinical application, it achieved a **recall of 88%** for the "Diseased" class. This means the model correctly identified 36 out of the 41 patients who actually had heart disease, making it a highly effective tool for minimizing false negatives and ensuring high-risk patients are not overlooked.

**5. Conclusion & Actionable Recommendations**

The analysis successfully identified several key clinical and demographic factors that are highly predictive of heart disease. The strong performance of the Tuned Random Forest model confirms that a data-driven approach can be effectively used for risk assessment.

Based on these findings, the following recommendations are proposed:

1. **Prioritize Asymptomatic Patients for Screening:** Clinicians should recognize that "Asymptomatic" chest pain is a high-risk indicator and should not be dismissed. Patients in this category, particularly males over 50, should be prioritized for further cardiovascular testing.
2. **Focus on Key Clinical Metrics:** Maximum Heart Rate (thalach) and ST Depression (oldpeak) should be considered primary indicators during stress tests and initial patient evaluations.
3. **Implement the Predictive Model:** The developed Random Forest model should be considered for deployment as a clinical decision-support tool to provide an objective, data-driven risk score for each patient, augmenting the clinician's professional judgment.