

Absolutely — here's a **complete example** of a **descriptive analytics report** using the **Cleveland Heart Disease dataset** from the UCI repository. This sample is structured using your requested sections and styled for clarity and professionalism.

Descriptive Analytics Report

Cleveland Heart Disease Dataset (UCI Repository)

Prepared by: [Your Name or Team] Date: July 2025 Version: 1.0

Executive Summary

This descriptive analytics report examines the Cleveland subset of the UCI Heart Disease dataset, summarizing the key demographic, clinical, and diagnostic characteristics of the patients. The goal is to explore patterns in heart disease occurrence across age, sex, chest pain types, and other medical indicators. This foundational understanding will inform future diagnostic and predictive studies, improve data quality awareness, and help identify potential high-risk populations.

Key observations include:

- Heart disease is more prevalent among males and individuals aged 55 and older.
 - Certain chest pain types and exercise-induced angina are more common in patients with heart disease.
 - There are notable missing values in cholesterol and resting ECG features, warranting attention.
-

Introduction

The Cleveland Heart Disease dataset is a widely used clinical dataset from the UCI Machine Learning Repository. It contains 303 patient records, each described by 14 features, including demographic information (e.g., age, sex), clinical test results (e.g., cholesterol, resting blood pressure), and a target variable representing the presence of heart disease.

This report presents a descriptive analysis of the dataset to uncover basic patterns, identify data quality issues, and provide a foundation for subsequent analytical efforts such as diagnostic and predictive modeling.

Data and Methodology

Data Source:

- **Dataset:** Cleveland subset of the UCI Heart Disease dataset
- **Records:** 303 patients
- **Target Variable:** num (0 = no heart disease, 1–4 = increasing severity)

Key Features Analyzed:

- Age, Sex, Chest Pain Type (**cp**)
- Resting Blood Pressure (**trestbps**), Cholesterol (**chol**)
- Fasting Blood Sugar (**fb**s), Resting ECG Results (**restecg**)
- Maximum Heart Rate Achieved (**thalach**), Exercise-Induced Angina (**exang**)
- ST Depression (**oldpeak**), Thalassemia (**thal**)

Methodology:

- Exploratory Data Analysis (EDA) using Pandas, Seaborn, and Matplotlib
 - Summary statistics, distribution plots, and cross-tabulations
 - Missing data identification and completeness checks
 - Data was not imputed, transformed, or modeled — only described
-

Analysis and Findings

1. Target Variable Distribution (Presence of Heart Disease)

- ~54% of patients have heart disease (target > 0)
- The most common severity class is 1 (mild disease)

2. Age Distribution

- Median age: 55 years
- Patients with heart disease tend to be older (median: 58 vs. 52)

3. Sex Distribution

- 68% of the dataset are male
- 84% of heart disease cases occur in males

4. Chest Pain Type (cp)

- Patients with typical angina (cp=0) are more likely to **not** have heart disease
- Patients with asymptomatic pain (cp=3) have the highest heart disease prevalence

5. Cholesterol (chol) and Resting BP (trestbps)

- Both features are right-skewed; extreme values exist
- Cholesterol has ~1% missing values

6. Exercise-Induced Angina (exang)

- Among patients with heart disease, 66% report angina during exercise
- Strong potential for use in early screening

7. Data Quality

- **thal** and **ca** have non-numeric missing values (“?”), affecting ~5% of rows
 - Feature scaling and data cleaning needed for future modeling
-

Recommendations

While descriptive analysis does not infer causality or make forecasts, the following next steps are recommended:

1. Data Cleaning:

- Replace or impute “?” values in **thal** and **ca**
- Standardize units and investigate outliers in cholesterol and blood pressure

2. Targeted Diagnostic Analysis:

- Explore the relationship between chest pain type and disease severity
- Examine interaction effects of `age`, `exang`, and `thalach` on heart disease presence

3. Gender-Specific Risk Profiling:

- Given the high male skew, consider stratifying future analysis by sex

4. Data Augmentation:

- Integrate other hospital sources or imaging data for richer context and completeness

5. Visualization for Stakeholders:

- Use interactive dashboards to communicate trends to clinicians and administrators

Conclusion

This descriptive analytics report provides an initial look into the Cleveland Heart Disease dataset, highlighting important population-level trends in age, sex, chest pain, and exercise response. These findings form a foundation for more advanced diagnostic analytics and predictive modeling, as well as operational decisions related to patient screening and data quality improvements.

Next steps should focus on deeper statistical analysis, careful preprocessing, and hypothesis-driven diagnostics to move toward actionable healthcare insights.

Would you like this exported to a markdown file, PDF, or Jupyter notebook template format to get you started?